Revenue Maximization in Probabilistic Single-Item Auctions via Signaling

Yuval Emek* Michal Feldman† Iftah Gamzu‡ Moshe Tennenholtz§

Abstract

Signaling is an important topic in the study of asymmetric information in economic settings. In particular, the transparency of information available to a seller in an auction setting is a question of major interest. We introduce the study of signaling when auctioning a probabilistic good whose actual instantiation is known to the auctioneer but not to the bidders. In particular, this can be used to model impressions selling in display advertising. Our study focuses on finding an optimal signaling scheme, captured by a segmentation of the possible goods into disjoint clusters. We show that while the problem is computationally hard, it possesses constant approximation for natural families of distribution functions over bidders valuations for the goods.

1 Introduction

Many auction settings encompass an inherent asymmetry of information between the auctioneer and the bidders competing for the auctioned item. We model this asymmetry by considering a framework termed a *probabilistic single-item auction*, in which n bidders participate in a second-price auction for a single item, which is chosen randomly from a set of m goods, according to a common knowledge probability distribution $p \in \Delta(m)$. In contrast to the bidders, who know only the probability distribution over the possible goods, the auctioneer knows its actual realization, and can use this informational superiority to increase the collected revenue.

In particular, the auctioneer may choose to reveal partial information to the bidders by means of a *signaling scheme*: the auctioneer a-priori declares a partition of the set of goods into clusters, and once an item is randomly chosen, she reveals the cluster that contains the chosen good to the bidders. The bidders then submit their bids, and the winner and payment are determined according to the second-price auction, namely, the winner is the bidder with the highest bid and the payment is the second highest bid. The auctioneer's goal — our goal in this paper — is to design a signaling scheme that maximizes her expected revenue.

A simple but crucial observation that facilities our analysis is that, similar to a "deterministic" second-price auction, here too, it is a dominant strategy for the bidders to reveal their true (expected) valuations for the signaled cluster. Therefore, the problem, termed *revenue maximization by signaling*, reduces to finding a partition of the goods into clusters in a way that maximizes the expected second highest bid (amounting to the expected revenue).

*ETH Zurich.
†Hebrew University of Jerusalem, and Microsoft Israel R&D Center.
‡Microsoft Israel R&D Center.
§Microsoft Israel R&D Center, and Technion-Israel Institute of Technology.

Interestingly, there are instances in which an effective signaling scheme provides a substantial improvement over the two *trivial* schemes, namely, the one that groups all the goods in a single cluster, thus revealing no information to the bidders, and the one that groups each good in its own cluster, thereby revealing the actual realization to the bidders. Consider for example the case in which there are n bidders and n good types, an item is chosen uniformly at uniform, and each bidder i is only interested in good i with valuation 1. If no information is revealed, then the expected revenue is $1/n$ as the expected valuation of each bidder is $1/n$. On the other extreme, if the actual realization is revealed, then no revenue is collected since for every realization, the second highest valuation is 0. However, if the goods are partitioned into clusters of size 2, then the expected revenue is $1/2$, providing a linear improvement over the best trivial scheme.

Since the bidders' valuations are rarely known to the auctioneer, we study the revenue maximization by signaling problem within a *Bayesian* setting in which the valuation of each bidder for each good is a random variable. Our results show that the problem is computationally hard, but can be approximated to within a constant factor for a natural family of these random variables. We also identify a special case that admits an optimal polynomial-time algorithm. A more detailed exposition of our results requires additional details, and thus, it is deferred to Section 2, where the formal setup is established.

An Application. One natural practical application that may be captured by our model is the billions of dollars display advertising market. In this market, publishers (such as MSN and Yahoo) attempt to maximize the revenue they collect from advertisers (such as Nike and Coca-Cola) for wisely targeting their ads at the right users. For example, an ad referring to the surfing lifestyle may be most valuable when targeted at a teenager from California; perhaps less valuable when targeted at older folks from Oregon; and even less valuable when targeted at people in areas that are distant from the ocean. Thus, each advertiser (a bidder in our model) has a different valuation for each impression type (a good in our model). While advertisers may know the distribution of users visiting a particular site, the publisher (the auctioneer in our model) usually has a much more accurate information about the site visitors.[1]

Various business models are carried out in the market for impressions [9]. In some cases, the publisher does not reveal any information to the bidders, and the advertisers bid on impressions, given their partial probabilistic view. In other situations, impressions are clustered into segments which are revealed to the bidders and sold, typically not through an auction mechanism. The current paper suggests that the publisher can improve her collected revenue by a well-designed signaling scheme that reveals partial information to the participants of an auction.

Related Work. There is a rich theory on markets admitting information asymmetry. In such markets, agents on one side have more or much better information than those on the other side. The foundation of this theory dates back to the 1970s when Akerlof, Spence, and Stiglitz analyzed markets with asymmetric information. (They received the 2001 Nobel prize for their work.) In particular, Akerlof [1] introduced the first formal analysis of markets in which sellers have more information than buyers regarding the quality of products. Spence [11, 12] demonstrated that in certain settings, well-informed agents can improve their outcome by signaling their private information to poorly-informed agents.

There is also a vast literature on the nature and effects of information revelation in auctions. One of the most fundamental results in auction theory, namely the 'Linkage Principle' of Milgrom and Weber [8], states that expected revenue of an auctioneer is enhanced when bidders are provided with more information. While this work advocates transparency in various markets, later work observed that such a transparency may not be

[1] In reality, the additional information about the visitor of a site is often handled by third party DSPs. For the sake of simplicity, we abstract away this distinction.

optimal in general (see, e.g., [10, 7, 3, 2]). Our work may be considered as a study of information revelation through an optimization lens, trying to maximize the expected revenue of an auctioneer by designing an effective information revelation scheme.

2 Model and Results

Probabilistic Single-Item Auctions. A *probabilistic single-item auction (PSIA)* $\mathcal{A}$ is formally depicted by the four-tuple

$$\mathcal{A} = \langle n, m, p, V \rangle \ ,$$

where $n \in \mathbb{Z}_{>0}$ stands for the number of *bidders*, $m \in \mathbb{Z}_{>0}$ stands for the number of distinct indivisible *goods*, $p \in \Delta(m)$ is a probability distribution over the goods, and $V \in \mathbb{R}_{\geq 0}^{n \times m}$ is a non-negative real matrix capturing the *valuation* $V_{i,j}$ of bidder i for good j. A single item $j \in [m]$ is chosen (by nature) according to the distribution p which is a common knowledge.

The auction is conducted according to the *second-price* rule: Each player i places her bid b_i and the chosen good j is sold to the bidder that placed the highest bid $\max_{i \in [n]}\{b_i\}$ (ties are broken arbitrarily) for the price of the second highest bid $\max2_{i \in [n]}\{b_i\}$.

Signaling Schemes. Although the bidders know the distribution p, they do not know its actual realization which is observed only by the *auctioneer*. In attempt to increase her expected *revenue*, the auctioneer may partially reveal the realization $j \in [m]$ of p to the bidders. This is performed via a *signaling scheme*: the auctioneer partitions the goods into pairwise disjoint clusters $C_1 \dot{\cup} \cdots \dot{\cup} C_k = [m]$ and reports this partition to the bidders; when good j is randomly chosen (happens with probability $p(j)$), the bidders are *signaled* the cluster C_ℓ that contains j. Each bidder then holds a "more accurate picture" of the chosen good j; this picture corresponds to the probability distribution p conditioned on the choice of some good in C_ℓ. In other words, the bidder knows that none of the goods in $[m] - C_\ell$ was chosen, but she still does not know exactly which good in C_ℓ was chosen.

It is important to point out that the signaling scheme $\{C_1, \ldots, C_k\}$ is decided by the auctioneer prior to nature's (random) choice of good j. The bidders know this partition, so when a bidder is signaled that *event* C_ℓ occurred, namely, that the chosen good is in cluster C_ℓ, she knows how to interpret it and calculate the conditional probability $\mathbb{P}(j \mid C_\ell) = p(j)/\mathbb{P}(C_\ell)$.

It is well known that in second-price "deterministic" single-item auctions, it is a dominant strategy for the bidders to be truthful, i.e., to bid their true valuations [13]. It turns out that this remains valid in probabilistic single-item auctions under signaling as well, as the following observation demonstrates. (The proof of this observation is deferred to the appendix.)

Observation 2.1. *For every $i \in [n]$ and $\ell \in [k]$, bidding $b_i(\ell) = \mathbb{E}_p[V_{i,j} \mid C_\ell]$ in response to the signal C_ℓ is a dominant strategy for bidder i.*

Optimization Problems. Consider some PSIA $\mathcal{A} = \langle n, m, p, V \rangle$ and signaling scheme $\mathcal{C} = \{C_1, \ldots, C_k\}$. In light of Observation 2.1, we subsequently assume that the bidders are indeed truthful, that is, bidder i bids $\mathbb{E}_p[V_{i,j} \mid C_\ell]$ in response to the signal C_ℓ. Therefore, the expected *revenue* of the auctioneer, denoted $\rho_{\mathcal{A}}(\mathcal{C})$, is given by

$$\rho_{\mathcal{A}}(\mathcal{C}) = \sum_{\ell \in [k]} \mathbb{P}\left(C_\ell\right) \cdot \max2_{i \in [n]} \left\{ \sum_{j \in C_\ell} \mathbb{P}(j \mid C_\ell) \cdot V_{i,j} \right\} \ .$$

When $\mathcal{A}$ is clear form the context, we may omit it and write simply $\rho(\mathcal{C})$. This raises the following combinatorial optimization problem, referred to as the *revenue maximization by signaling (RMS) problem*: given a PSIA $\mathcal{A}$, construct the signaling scheme $\mathcal{C}$ that maximizes $\rho_{\mathcal{A}}(\mathcal{C})$. Note that we do not impose any requirements on the number k of clusters in the signaling scheme $\mathcal{C}$; in particular, we consider the two extreme cases of the signaling scheme that packs all goods in one cluster ($k = 1$), thus not revealing any information to the bidders, and the signaling scheme that packs each good in its own singleton cluster ($k = m$), thus revealing the complete information to the bidders.

Notice that the expected revenue of the auctioneer can be rewritten as

$$\begin{aligned}\rho(\mathcal{C}) &= \sum_{\ell\in[k]} \mathbb{P}\left(C_\ell\right) \cdot \max2_{i\in[n]} \left\{ \sum_{j\in C_\ell} \frac{p(j)}{\mathbb{P}\left(C_\ell\right)} \cdot V_{i,j} \right\} \\ &= \sum_{\ell\in[k]} \max2_{i\in[n]} \left\{ \sum_{j\in C_\ell} p(j) \cdot V_{i,j} \right\} \\ &= \sum_{\ell\in[k]} \max2_{i\in[n]} \left\{ \sum_{j\in C_\ell} \Psi_{i,j} \right\} , \qquad (1)\end{aligned}$$

where $\Psi_{i,j} = p(j) \cdot V_{i,j}$. The aforementioned RMS problem can therefore be reduced to the *simplified revenue maximization by signaling (SRMS) problem*: Given a matrix $\Psi \in \mathbb{R}^{n\times m}_{\geq 0}$, construct the partition $\mathcal{C} = \{C_1, \dots, C_k\}$ of $[m]$ that maximizes $\rho(\mathcal{C}) = \sum_{\ell\in[k]} \max2_{i\in[n]}\{\sum_{j\in C_\ell} \Psi_{i,j}\}$.

A Bayesian Setting. Up until now, we implicitly assumed that the valuations of the bidders are known to the auctioneer.[2] However, in many practical scenarios the auctioneer does not know the exact valuation of each bidder. To tackle this obstacle, we assume a *Bayesian* setting, treating the state of knowledge that the auctioneer holds on the bidders' valuations in a probabilistic manner. This is captured in our model by considering each valuation $V_{i,j}$ to be a non-negative real random variable (rather than a fixed value), which means that V is a random matrix. Consequently, $\Psi_{i,j}$ is also considered to be a non-negative real random variable, and Ψ a random matrix. It is important to point out that unless stated otherwise, we do not assume that the random variables $V_{i,j}$ are independent.

The expected revenue of the auctioneer from the signaling scheme $\mathcal{C}$ is now defined to be

$$\begin{aligned}\rho(\mathcal{C}) &= \mathbb{E}_{p,V}\left[\sum_{\ell\in[k]} \mathbb{P}\left(C_\ell\right) \cdot \max2_{i\in[n]} \left\{ \sum_{j\in C_\ell} \mathbb{P}(j \mid C_\ell) \cdot V_{i,j} \right\} \right] \\ &= \mathbb{E}_{\Psi}\left[\sum_{\ell\in[k]} \max2_{i\in[n]} \left\{ \sum_{j\in C_\ell} \Psi_{i,j} \right\} \right] ,\end{aligned}$$

where the last equation follows from the same line of arguments that was used to establish (1).

Given a PSIA $\mathcal{A} = \langle n, m, p, V\rangle$ and a signaling scheme $\mathcal{C}$, let $\mathcal{C}(\mathcal{A})$ be the random variable that stands for the revenue of the auctioneer from $\mathcal{C}$, so $\mathbb{E}_{p,V}[\mathcal{C}(\mathcal{A})] = \rho_{\mathcal{A}}(\mathcal{C})$. When the parameters n, m, and p are clear from the context and we wish to emphasize that $\mathcal{C}(\mathcal{A})$ refers to the (possibly random) valuation matrix V, we

[2] In some sense, we also assumed that the valuations of each bidder are known to the other bidders. However, Observation 2.1 implies that this does not matter: a bidder is better off bidding its true (expected) valuation regardless of the strategies of the other bidders.

may write $\mathcal{C}(V)$. In the context of the SRMS problem, we will replace $\mathcal{C}(V)$ with $\mathcal{C}(\Psi)$. The notation $\texttt{Opt}$ will denote an optimal signaling scheme. The terms bidder and row and the terms good and column are used interchangeably.

Our Results. The RMS problem is shown to be strongly NP-hard, even when the valuations of the bidders are fixed, i.e., when the valuation matrix is not random (Section 5). Thus, the problem does not admit an FPTAS, unless P=NP. On the positive side, for a natural family of valuation distributions, including the case in which each valuation is uniformly distributed in some interval (the intervals are not necessarily identical and the random variables need not be independent), a constant-factor approximation[3] is established. It is interesting to point out that the approximation ratio of our algorithm remains constant even when compared to the expected performance of an optimal algorithm that knows the realization of the random valuation matrix (Section 3). Finally, we develop an optimal polynomial-time algorithm for the special case of the SRMS problem, when the valuation matrix is binary, that is, each $\Psi_{i,j} \in \{0, 1\}$ (Section 4).

3 Approximation Algorithm for RMS

In this section we consider a family of PSIAs $\mathcal{A} = \langle n, m, p, V \rangle$ in which each valuation $V_{i,j}$ is uniformly distributed in the interval $[a_{i,j}, b_{i,j}]$ for some $0 \leq a_{i,j} \leq b_{i,j}$, and establish a constant approximation for RMS instances in this family. More generally, we design an efficient algorithm $\texttt{Alg}$ with the following performance guarantee.

Theorem 3.1. *Consider some PSIA $\mathcal{A} = \langle n, m, p, V \rangle$ and real $c \geq 1$, and assume that $V_{i,j} \leq c \cdot \mathbb{E}[V_{i,j}]$ with probability 1 for every $i \in [n]$ and $j \in [m]$. Then*

$$\mathbb{E}_{p,V}\left[\texttt{Alg}(V)\right] \;\geq\; \Omega(1/c) \cdot \mathbb{E}_p\left[\int_{v \in \mathbb{R}_{\geq 0}^{n \times m}} \mathbb{P}(V = v) \cdot \texttt{Opt}(v)\, dv\right] .$$

In other words, $\texttt{Alg}$ is an $O(c)$ approximation algorithm even when compared to the expected performance of an optimal algorithm that knows the realization of the random valuation matrix V.

As the uniform distribution in the interval $[a_{i,j}, b_{i,j}]$ for some $0 \leq a_{i,j} \leq b_{i,j}$ meets the condition of Theorem 3.1 for $c = 2$, we obtain a constant approximation for uniform valuations. Note that apart from this condition, we do not impose any other assumptions on the random variables $V_{i,j}$. In particular, these random variables may exhibit complicated dependencies.

Our approximation approach is carried out in three steps. First, we apply the reduction of RMS to SRMS, introducing the random matrix $\Psi \in \Delta(\mathbb{R}_{\geq 0}^{n \times m})$ (see Section 2). Then, in Section 3.1, we develop a constant approximation algorithm $\texttt{Alg}^{\text{fixed}}$ for the SRMS problem on fixed valued matrices (corresponds to the auctioneer knowing the exact valuation of each bidder for every good). The desired approximation is established by invoking $\texttt{Alg}^{\text{fixed}}$ on the (fixed valued) matrix $\bar{\Psi}$ obtained from Ψ by setting $\bar{\Psi}_{i,j} = \mathbb{E}[\Psi_{i,j}]$ for every $i \in [n]$ and $j \in [m]$. This is a consequence of the following proposition.

Proposition 3.2. *Suppose that $\Psi_{i,j} \leq c \cdot \bar{\Psi}_{i,j}$ with probability 1 for every $i \in [n]$ and $j \in [m]$. Then, $\int_{\psi \in \mathbb{R}_{\geq 0}^{n \times m}} \mathbb{P}(\Psi = \psi) \cdot \texttt{Opt}(\psi)\, d\psi \leq c \cdot \texttt{Opt}(\bar{\Psi})$.*

[3] In this version of the paper, we do not attempt to optimize the constants.

Proof. Consider some $\psi \in \mathbb{R}_{\geq 0}^{n \times m}$ such that $\mathbb{P}(\Psi = \psi) > 0$. By definition, $\psi_{i,j} \leq c \cdot \bar{\Psi}_{i,j}$ for every $i \in [n]$ and $j \in [m]$. Now, suppose that $\mathcal{C}^*$ is the partition that realizes $\texttt{Opt}(\psi)$. Then,

$$\texttt{Opt}(\psi) \;=\; \mathcal{C}^*(\psi) \;\leq\; c \cdot \mathcal{C}^*(\bar{\Psi}) \;\leq\; c \cdot \texttt{Opt}(\bar{\Psi})$$

(recall that $\mathcal{C}^*(\psi)$ and $\mathcal{C}^*(\bar{\Psi})$ stand for the revenue obtained from the partition $\mathcal{C}^*$ in the matrices ψ and $\bar{\Psi}$, respectively). The assertion follows. □

3.1 Constant Approximation for Fixed Valuations

Consider some (fixed valued) input matrix $\Psi \in \mathbb{R}_{\geq 0}^{n \times m}$ for the SRMS problem. For every $j \in [m]$, let $\eta(j)$ denote some bidder $i \in [n]$ that maximizes $\Psi_{i,j}$. In addition, for every $i \in [n]$, fix $G_i = \{j \in [m] : \eta(j) = i\}$ and $\gamma_i = \sum_{j \in G_i} \Psi_{i,j}$. Our algorithm, denoted $\texttt{Alg}^{\text{fixed}}$, works as follows.

The matching step: As long as there exist $i, i' \in [n]$, $i \neq i'$, such that $\gamma_i \leq \gamma_{i'} \leq 2\gamma_i$, form a new cluster S that consists of all the goods in $G_i \cup G_{i'}$ and remove rows i and i' and the columns in S from Ψ.

Let T be the set of goods that were clustered in the matching step. Assume without loss of generality that the remaining rows in Ψ are rows $1, \dots, n'$ and that $\gamma_1 \geq \cdots \geq \gamma_{n'}$. We refer to the goods in $[m] - T$ as the *unmatched* goods.

Observation 3.3. *The unmatched goods satisfy $\gamma_i > 2\gamma_{i+1}$ for every $1 \leq i \leq n' - 1$.*

The branching step: The algorithm continues by comparing between two alternatives and selecting the alternative that yields a higher revenue:

- Form a singleton cluster for every (unmatched) good $j \in G_1$ (all other unmatched goods can be ignored).
- Form a single cluster S that consists of all the (unmatched) goods in $G_1 \cup G_2$ (all other unmatched goods can be ignored).

We now turn to analyze the performance of our algorithm.

Lemma 3.4. $\texttt{Alg}^{\textit{fixed}}$ *guarantees a constant approximation ratio.*

Proof. Consider some partition $\mathcal{C}$ of $[m]$ ($\mathcal{C}$ will take the role of the partition returned by either $\texttt{Opt}$ or $\texttt{Alg}^{\text{fixed}}$). Let j be some good in $[m]$ and let $C \in \mathcal{C}$ be the cluster of $\mathcal{C}$ that contains j. Let $i(C)$ be the bidder that realizes $\max2_{i \in [n]}\{\sum_{j' \in C} \Psi_{i,j'}\}$, where we break ties arbitrarily, but in a consistent manner. We refer to $\Psi_{i(C),j}$ as the *revenue* obtained by the partition $\mathcal{C}$ from good j, denoted $\mathcal{C}(j)$ (which will be replaced by either $\texttt{Opt}(j)$ or $\texttt{Alg}^{\text{fixed}}(j)$). Given some subset $U \subseteq [m]$, we define

$$\begin{aligned}
\texttt{Opt}(U) &= \sum_{j \in U} \texttt{Opt}(j) , \\
\texttt{Alg}^{\text{fixed}}(U) &= \sum_{j \in U} \texttt{Alg}^{\text{fixed}}(j) , \\
\varphi(U) &= \sum_{j \in U} \max_{i \in [n]}\{\Psi_{i,j}\} \text{ , and} \\
\varphi_2(U) &= \sum_{j \in U} \max2_{i \in [n]}\{\Psi_{i,j}\} .
\end{aligned}$$

By definition, we know that $\varphi(U) \geq \max\{\texttt{Opt}(U), \texttt{Alg}^{\text{fixed}}(U), \varphi_2(U)\}$.

Since each cluster formed by $\texttt{Alg}^{\text{fixed}}$ in the matching step matches the goods in G_i with the goods in $G_{i'}$ for some $i, i' \in [n]$ such that $\gamma_i \leq \gamma_{i'} \leq 2\gamma_i$, it follows that

$$\texttt{Alg}^{\text{fixed}}(T) \;\geq\; \frac{\varphi(T)}{3} \;\geq\; \frac{\texttt{Opt}(T)}{3}\ . \tag{2}$$

On the other hand, notice that the branching step guarantees that

$$\texttt{Alg}^{\text{fixed}}(G_1 \cup G_2) \;\geq\; \max\left\{\varphi_2(G_1), \varphi(G_2)\right\}\ . \tag{3}$$

Observation 3.3 implies that $2 \cdot \varphi(G_2) \geq \varphi(G_2 \cup \cdots \cup G_{n'}) \geq \texttt{Opt}(G_2 \cup \cdots \cup G_{n'})$, and thus,

$$\texttt{Opt}(G_2 \cup \cdots \cup G_{n'}) \;\leq\; 2 \cdot \texttt{Alg}^{\text{fixed}}(G_1 \cup G_2)\ . \tag{4}$$

Next, we turn to bound the revenue $\texttt{Opt}(G_1)$ obtained by $\texttt{Opt}$ from the goods in G_1. To that end, it will be convenient to identify the collection S of goods in G_1 that $\texttt{Opt}$ assigns to clusters dominated by bidder 1, i.e., clusters C such that $\sum_{j \in C} \Psi_{1,j} \geq \sum_{j \in C} \Psi_{i,j}$ for every $i \in [n]$. By definition, the revenue obtained by $\texttt{Opt}$ from the goods in S is $\texttt{Opt}(S) \leq \varphi_2(S) \leq \varphi_2(G_1)$. The key observation now is that the revenue obtained by $\texttt{Opt}$ from the goods in $G_1 - S$ cannot be larger than $\varphi(G_2 \cup \cdots \cup G_{n'} \cup T)$. Therefore,

$$\begin{aligned}
\texttt{Opt}(G_1) \;&\leq\; \varphi(G_2 \cup \cdots \cup G_{n'} \cup T) + \varphi_2(G_1) \\
&=\; \varphi(G_2 \cup \cdots \cup G_{n'}) + \varphi(T) + \varphi_2(G_1) \\
&\leq\; 2 \cdot \varphi(G_2) + \varphi(T) + \varphi_2(G_1) \qquad (5)\\
&\leq\; 3 \cdot \texttt{Alg}^{\text{fixed}}(G_1 \cup G_2) + 3 \cdot \texttt{Alg}^{\text{fixed}}(T)\ , \qquad (6)
\end{aligned}$$

where, (5) is due to Observation 3.3 and (6) follows from (2) and (3). By combining (2), (4), and (6), we conclude that

$$\begin{aligned}
\texttt{Opt}([m]) \;&=\; \texttt{Opt}(G_1) + \texttt{Opt}(G_2 \cup \cdots \cup G_{n'}) + \texttt{Opt}(T) \\
&\leq\; 3 \cdot \texttt{Alg}^{\text{fixed}}(G_1 \cup G_2) + 3 \cdot \texttt{Alg}^{\text{fixed}}(T) + 2 \cdot \texttt{Alg}^{\text{fixed}}(G_1 \cup G_2) + 3 \cdot \texttt{Alg}^{\text{fixed}}(T) \\
&\leq\; 6 \cdot \texttt{Alg}^{\text{fixed}}(G_1 \cup G_2) + 6 \cdot \texttt{Alg}^{\text{fixed}}(T) \\
&=\; 6 \cdot \texttt{Alg}^{\text{fixed}}([m])\ .
\end{aligned}$$

The assertion follows. □

4 Optimal Algorithm for Binary Matrices

In this section we consider a special case of the SRMS problem, in which the entries of Ψ are binary, i.e., $\Psi_{i,j} \in \{0, 1\}$ for every $i \in [n]$ and $j \in [m]$. We show that an optimal partition for this special case can be constructed in polynomial time. Our algorithm works in two steps:

1. Form a singleton cluster $\{j\}$ for every column $j \in [m]$ with at least two 1-entries, and remove this column from the matrix.
2. Construct an undirected graph $G = (V_G, E_G)$, where V_G is identified with the remaining columns of the matrix, and $(j, j') \in E_G$ if and only if there exist $i, i' \in [n]$, $i \neq i'$, such that $\Psi_{i,j} = \Psi_{i',j'} = 1$. In other words, the vertices corresponding to the columns of the matrix Ψ; columns j and j' are connected by an edge if and only if the restriction of Ψ to these two columns admits a generalized diagonal. Find a maximum matching M in G, and form a cluster $\{j, j'\}$ for every $(j, j') \in M$.

We would like to point out that our algorithm can be implemented to run in linear time. In particular, the matching step can be implemented to run in linear time due to some structural properties of our instance. Notice that after the first step, each column has exactly one non-zero entry. Suppose that m' columns were left in the matrix after that step, and let $G_i = \{j \in [m'] : \Psi_{i,j} = 1\}$ be the set of columns that can be associated with row i. Let us assume without loss of generality that $|G_1| \geq |G_2| \geq \cdots \geq |G_n|$. Consider the ordering that places all the columns of G_1 first, the columns of G_2 second, and so on. One can verify that matching each column $r \in \{1, \ldots, \lfloor m'/2 \rfloor\}$ in the ordering with column $r + \lceil m'/2 \rceil$ in the ordering yields a maximum matching of columns.

We now turn to prove that our algorithm returns an optimal partition. In general, our algorithm can output several different partitions (the graph G may have several different maximum matchings), but they all produce the same revenue. Among all optimal partitions for Ψ, let $\mathcal{C} = \{C_1, \ldots, C_k\}$ be the one that maximizes the number of clusters k. We will show that $\mathcal{C}$ can be output by our algorithm, which means that our algorithm outputs an optimal partition. Our first step is to prove that a column with two 1-entries must form a singleton cluster in $\mathcal{C}$.

Proposition 4.1. *If there exists some $j \in [m]$ and $i, i' \in [n]$, $i \neq i'$, such that $\Psi_{i,j} = \Psi_{i',j} = 1$, then $\{j\}$ is a singleton cluster in $\mathcal{C}$.*

Proof. Suppose towards deriving contradiction that column j belongs to some cluster C_ℓ of size $|C_\ell| > 1$. Then by removing j from C_ℓ, forming a new singleton cluster $\{j\}$, the revenue obtained from C_ℓ (which does not turn empty) decreases by at most 1, but this is compensated by the revenue obtained from the new singleton cluster. Therefore, we get a new partition whose revenue is at least as good as that of $\mathcal{C}$, but whose size is larger, in contradiction to the maximality of $\mathcal{C}$. □

Next,we show the clusters in $\mathcal{C}$ are of size at most 2.

Proposition 4.2. *The partition $\mathcal{C}$ satisfies $|C_\ell| \leq 2$ for every $\ell \in [k]$.*

Proof. Suppose towards deriving contradiction that there exists some cluster $C_\ell \in \mathcal{C}$ of size $|C_\ell| > 2$. We show that C_ℓ can be (re)partitioned into (sub)clusters without decreasing the revenue, thus contradicting the maximality of $\mathcal{C}$. For this purpose, let us assume without loss of generality that $\sum_{j \in C_\ell} \Psi_{1,j} \geq \sum_{j \in C_\ell} \Psi_{2,j} \geq \sum_{j \in C_\ell} \Psi_{i,j}$ for every $i \in \{3, \ldots, n\}$. Clearly, there cannot exist any column $j \in C_\ell$ such that $\Psi_{1,j} = \Psi_{2,j} = 0$ as such a column can be removed from C_ℓ, forming a new singleton cluster, without decreasing the revenue obtained from C_ℓ. We may assume, due to Proposition 4.1, that $\Psi_{1,j} = 0$ for every column $j \in C_\ell$ such that $\Psi_{2,j} = 1$, and vice versa. Since row 1 has at least as many 1-entries as row 2, we can identify two columns $j, j' \in C_\ell$, $j \neq j'$, such that $\Psi_{1,j} = \Psi_{2,j'} = 1$ and $\Psi_{1,j'} = \Psi_{2,j} = 0$. By removing these two columns from C_ℓ and forming a new cluster $\{j, j'\}$, the revenue obtained from C_ℓ (which does not turn empty as originally $|C_\ell| > 2$) decreases by at most 1, but this is compensated by the revenue obtained from the new cluster. □

The analysis is completed by observing that the non-singleton clusters in $\mathcal{C}$ correspond to edges in the graph G, so their overall structure corresponds to a matching in G. This matching must be maximum as $\mathcal{C}$ is assumed to be optimal.

5 Hardness of RMS

In this section, we demonstrate that the SRMS problem, and hence also the RMS problem, is NP-hard in the strong sense. To this end, we exhibit a reduction from 3-Partition, as defined below, to the decision variant

SRMS.

Definition 1 (3-Partition)**.** Given a multiset $S = \{a_1, a_2, \dots, a_n\}$ of $n = 3m$ positive integers, determine if S can be partitioned into m subsets $S_1, S_2, \dots, S_m$ such that sum of the numbers in each subset is equal.

Theorem 5.1 ([6, 4])**.** *3-Partition is strongly NP-complete, namely, NP-complete even when the integers $\{a_1, a_2, \dots, a_n\}$ are bounded by a polynomial in n. Let $\Lambda = \sum_{i=1}^{n} a_i$ and $\lambda = \Lambda/m$. Then the problem remains strongly NP-complete even when each integer a_i satisfies $a_i \in (\lambda/4, \lambda/2)$ which means that each subset S_i must be of size exactly $|S_i| = 3$.*

The main result of this section is the following.

Theorem 5.2. *Given a (fixed valued) matrix $\Psi \in \mathbb{Z}_{\geq 0}^{n \times m}$ and an integer α, it is strongly NP-hard to determine if the SRMS problem on input Ψ admits a partition whose revenue is at least α.*

Proof. Let $S = \{a_1, a_2, \dots, a_n\}$ be an instance of 3-Partition with $n = 3m$. Fix $\Lambda = \sum_{i=1}^{n} a_i$ and $\lambda = \Lambda/m$. We construct an instance of the SRMS problem by setting $\alpha = \Lambda$ and setting the matrix $\Psi \in \mathbb{Z}_{\geq 0}^{(m+1)\times(m+n)}$ to be

$$\Psi = \begin{pmatrix} \lambda & 0 & 0 & 0 & \cdots & 0 \\ 0 & \ddots & 0 & \vdots & \ddots & \vdots \\ 0 & 0 & \lambda & 0 & \cdots & 0 \\ 0 & \cdots & 0 & a_1 & \cdots & a_n \end{pmatrix}.$$

Completeness. Let $S_1, S_2, \dots, S_m$ be a feasible 3-Partition of the instance S. One can easily validate that the cluster C_i which consists of good i and the goods that correspond to the integers in S_i obtains a revenue of exactly λ. Accordingly, the signaling scheme $\{C_1, \dots, C_m\}$ collects a revenue of Λ.

Soundness. Suppose that there is a signaling scheme $\mathcal{C}$ that collects a revenue of Λ. Note that a cluster in $\mathcal{C}$ that does not contain any of the goods $\{1, \dots, m\}$ cannot obtain a positive revenue. Since no cluster can obtain a revenue larger than λ, and since $\Lambda = m \cdot \lambda$, it follows that each cluster in $\mathcal{C}$ contains exactly one of the goods $\{1, \dots, m\}$ and a non-empty subset of the goods $\{m+1, \dots, n\}$. In particular, the revenue obtained from each of the m clusters in $\mathcal{C}$ is exactly λ. Therefore, the collection $\{S(C) \mid C \in \mathcal{C}\}$, where $S(C)$ is the subset of positive integers a_i appearing in the last row of the goods in cluster C, is a feasible solution for the given 3-Partition instance. □

Corollary 5.3 follows from a well-known result of Garey and Johnson [5] regarding the intractability of strong NP-hard problems.

Corollary 5.3. *The SRMS problem does not admit an FPTAS unless P=NP.*

6 Conclusion

This work proposes a new approach for revenue maximization in auction settings with asymmetric information between the auctioneer and the bidders. We formulate the revenue maximization by signaling problem, which pursues for the optimal informational structure that should be revealed to the bidders. Our results show that while the problem is computationally hard, it can be approximated within constant factors for natural families of Bayesian valuations. This suggests several avenues for future directions: First, while our strong NP-hardness result precludes the existence of an FPTAS, the existence of a PTAS remains open. Second, we wish to further our positive approximability results to more families of valuation matrix distributions. For example, our proposed scheme works when the valuations are random variables which are sufficiently concentrated around the expectation. It is not difficult to show that the same scheme works when these random variables are independent with critical ratios[4] at most $O(1/m)$. We conjecture that this can be extended: with a small modification to the approximation approach, a constant approximation ratio can be guaranteed as long as the critical ratios are $O(1)$.

References

[1] G. A. Akerlof. The market for 'Lemons': quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3):488-500, 1970.

[2] Y. Feinberg and M. Tennenholtz. Anonymous bidding and revenue maximization. *Topics in Theoretical Economics*, 5(1), 2005.

[3] T. Foucault and S. Lovo, Linkage principle, multi-dimensional signals and blind auctions. Technical Report, HEC School of Management, 2003.

[4] Michael R. Garey and David S. Johnson. Complexity results for multiprocessor scheduling under resource constraints. *SICOMP*, 4(4):397–411, 1975.

[5] Michael R. Garey and David S. Johnson. Strong NP-completeness results: motivation, examples, and implications. *J. ACM*, 25(3):499–508, 1978.

[6] Michael R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, 1979.

[7] V. Krishna. *Auction Theory*, Academic Press, 2002.

[8] P. Milgrom and R. J. Weber. A theory of auctions and competitive bidding. *Econometrica*, 50:1089–1122, 1982.

[9] S. Muthukrishnan. Ad exchanges: Research issues. In *WINE*, pages 1–12, 2009.

[10] M. Perry and P.J. Reny. On the failure of the linkage principle in multi-unit auctions. *Econometrica*, 67(4):895–900, 1999.

[11] M. Spence. Job market signaling. *The Quarterly Journal of Economics*, 87(3):355-374, 1973.

[12] M. Spence. Signaling in retrospect and the informational structure of markets. *The American Economic Review*, 92(3):434-459, 2002.

[13] W. Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *Journal of Finance*, 16:8–37, 1961.

[4] The critical ratio of a random variable is the ratio of the variance to the square of the expectation.

APPENDIX

Proof of Observation 2.1. Consider the complete information *ex-ante* game defined by setting the strategy space of each bidder $i \in [n]$ to be the collection of all possible functions $b_i : [k] \to \mathbb{R}_{\geq 0}$ and the utility of bidder i from strategy profile $b = (b_1, \ldots, b_n)$ to be the expected utility of bidder i in the PSIA $\mathcal{A}$ assuming that each bidder adheres to b. Fix some (arbitrary) strategies $b_{i'} : [k] \to \mathbb{R}_{\geq 0}$ for all bidders $i' \neq i$ and consider the strategy b_i of bidder i that bids

$$b_i(\ell) = \mathbb{E}_p \left[V_{i,j} \mid C_\ell \right] = \sum_{j \in C_\ell} \mathbb{P}(j \mid C_\ell) \cdot V_{i,j}$$

in response to each signal C_ℓ, $\ell \in [k]$.

Fix some signal C_ℓ. From bidder i's point of view, the expected valuation of the chosen good is $b_i(\ell)$, whereas each other bidder $i' \neq i$, bids $b_{i'}(\ell)$. If bidder i does not win the chosen good, which happens only if $\max_{i' \neq i}\{b_{i'}(\ell)\} \geq b_i(\ell)$, then her utility in the ex-ante game is 0. This can be changed only if bidder i increases her bid so that it exceeds $\max_{i' \neq i}\{b_{i'}(\ell)\}$, but this imposes a negative utility on i. So, assume that $\max_{i' \neq i}\{b_{i'}(\ell)\} \leq b_i(\ell)$ and bidder i does win the chosen good. By the definition of the second-price rule, the utility of i must be non-negative. Clearly, bidder i has no incentive to increase her bid. Decreasing her bid does not change her utility as long as it still exceeds $\max_{i' \neq i}\{b_{i'}(\ell)\}$; decreasing her bid further resets her utility to zero. The assertion follows. □